Hierarchical Bayesian modeling of species richness: impacts of including spatial autocorrelation

Matthew Krachey, Department of Zoology, NC State University, Raleigh, NC
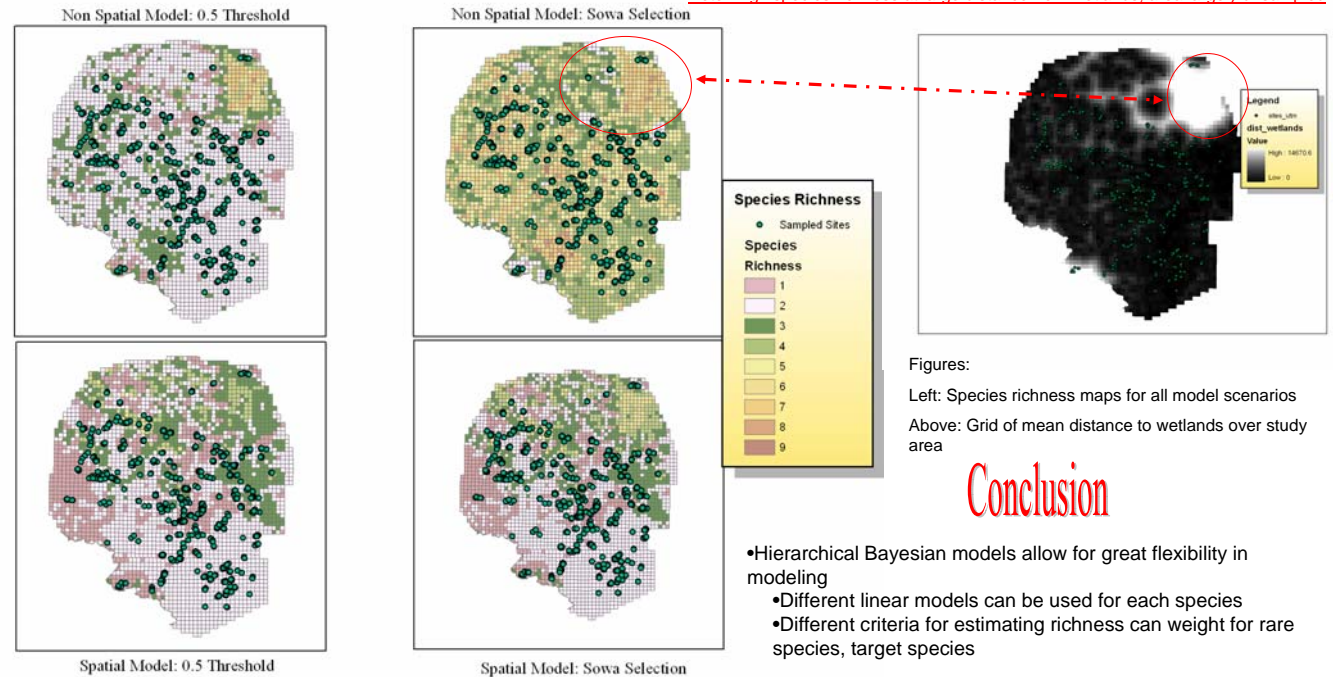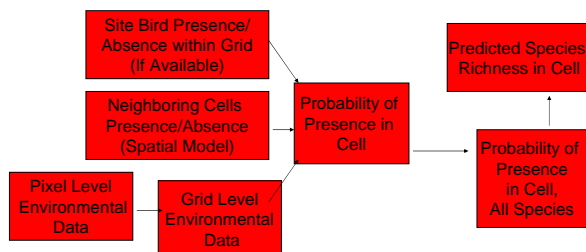
Introduction:

Conservation plans often seek to make reserves to maximize species diversity in a region. Selection of appropriate sites is often complicated, modeling approaches typically do not allow for heterogeneity of species or of locations. Here, a **Bayesian logistic regression** for **species richness** is applied allowing for **each species** to be modeled **individually**, and richness metrics are applied once probabilities of presence are estimated following Gelfand et al, 2005. **Random effects** for location are also estimated, allowing for freedom to note patterns unobserved in the linear predictors.

Methods (Dataset)

- Species richness of 15 warbler species was modeled from Michigan dataset provided by Ed Laurent
  - There were 433 observations used, 2987 grid cells
- Linear model component uses covariates averaged over grid cells
  - Used mean distance to wetlands, lakes, and streams, as well as proportion of grid comprised of aspen, northern hardwoods, and water pixels in the landcover dataset

Methods (Model)

- **Non-spatial model** uses "simple" hierarchical Bayesian logistic regression
  - Key parameter is the probability of presence in each grid level
- **Spatial model** adds a random effect in the logistic equation for each grid cell
  - Random effects follow a conditional autoregressive model
  - Random effects used to increase probability of presence if a neighboring cell has the species, and decrease the probability if a neighboring cell lacks the species

- **Species richness** defined by summing the expected presence of each species for each grid cell
  - **0.5 Threshold:** Present (presence=1) in grid if probability of presence is greater than 0.5
  - **Sowa threshold:** Present if probability of presence is greater than 0.5*(Global maximum presence probability for that species), increasing the probability of including rare species
- Used SAS, R, ArcGIS, WinBUGS, Excel

Site Bird Presence/Absence within Grid (If Available)

Neighboring Cells Presence/Absence (Spatial Model)

Pixel Level Environmental Data

Grid Level Environmental Data

Probability of Presence in Cell

Probability of Presence in Cell, All Species

Predicted Species Richness in Cell

Non Spatial Model: 0.5 Threshold

Non Spatial Model: Sowa Selection

Spatial Model: 0.5 Threshold

Spatial Model: Sowa Selection

Species Richness
- Sampled Sites
- **Species Richness**
  - 1
  - 2
  - 3
  - 4
  - 5
  - 6
  - 7
  - 8
  - 9

Figures:

Left: Species richness maps for all model scenarios

Above: Grid of mean distance to wetlands over study area

Results

- **Non-spatial model:**
  - Results in higher expected species richness across the majority of the landscape relative to the spatial model
  - Highly influenced by the extreme **high values of distance to wetlands** (see figure), **largely unsampled**
  - Sowa threshold left huge impact on landscape, implying that large numbers of species have very low maximum predicted presence

- **Spatial model:**
  - Results in relatively low abundances, relative to non-spatial model
  - Tighter clustering in some areas
  - Sowa threshold made little impact, implying the model predicted much higher presences for maximum presence, but many low values as well

- **Wetlands:**
  - Large impact when large distance magnitude in NE corner
  - **Sign on the mean coefficient value changes between models (below)**

	Mean Beta Coefficient Value Across All Species					
	% North. Hardwoods	%Aspen	Dist. Streams	**Dist. Wetlands**	Dist. Lakes	% Water
Non-spatial Model	-0.775	-0.021	-0.0012	**0.00013**	-0.000151	1.187
Spatial Model	-0.127	-0.0234	-0.00156	**-0.000124**	-0.000197	0.107

Conclusion

- Hierarchical Bayesian models allow for great flexibility in modeling
  - Different linear models can be used for each species
  - Different criteria for estimating richness can weight for rare species, target species

- **Non-spatial model** is very sensitive to choice of input model
  - Intensity of sampling design may also influence prediction

- **Spatial model**
  - Relatively robust, could provide management with more information to **find target reserve areas**
  - Seems to be more capable of providing strong weighting of presence probabilities
  - Distribution of the random effects parameters may provide additional insight

Future Research

Assessment
- More robust model selection for each species should be done if using this method
- Each species may be evaluated using cross-validation metrics, Kappa may be appropriate

- Incorporation of species interactions
- Investigation in cells with very high random effects values may be insightful
  - Differ from the linear model, may suggest unknown process important to species
  - May suggest connectivity very important

Literature cited:
Alan E. Gelfand, Alexandra M. Schmidt, Shanshan Wu, John A. Silander Jr, Andrew Latimer, Anthony G. Rebelo (2005) Modelling species diversity through species level hierarchical modelling Journal of the Royal Statistical Society: Series C (Applied Statistics) 54 (1), 120.